

Multiple class I and class II *Haemophilus ducreyi* strains cause cutaneous ulcers in children on an endemic island

*Jacob C. Grant¹, *Camila González-Beiras², Kristen M. Amick³, Kate R. Fortney¹, Dharanesh Gangaiah¹, Tricia L. Humphreys³, Oriol Mitjà^{4,5,6}, Ana Abecasis⁷, and Stanley M. Spinola^{1,8,9}

¹Departments of Microbiology and Immunology, ⁸Medicine, ⁹Pathology and Laboratory Medicine, Indiana University School of Medicine, Indiana University, Indianapolis, Indiana, USA; ²Global Public Health PhD Programme, Nova University of Lisbon, Lisbon, Portugal; ³Department of Biology, Allegheny College, Meadville, PA, USA; ⁴Lihir Medical Centre–International SOS, Newcrest Mining, Lihir Island, Papua New Guinea; ⁵Barcelona Institute for Global Health, Hospital Clinic, University of Barcelona, Barcelona, Spain; ⁶School of Medicine and Health Sciences, University of Papua New Guinea, Port Moresby, Papua New Guinea; and ⁷Global Health and Tropical Medicine, Institute of Hygiene and Tropical Medicine, Lisbon, Portugal

* These authors contributed equally to this work

Corresponding Author: Dr. Stanley M. Spinola, 635 Barnhill Drive, Room MS420, Indianapolis, IN 46202-5124 (sspinola@iu.edu) Tel: 317-274-0506, Fax: 317-278-3331.

Summary

H. ducreyi is as a major cause of skin ulcers in the tropics. On an endemic island, multiple strains of *H. ducreyi* cause infection, co-infections are common, and mass treatment with azithromycin did not exert selection pressure on the organism.

Running title: Epidemiology of *H. ducreyi* skin ulcers

This is the author's manuscript of the article published in final edited form as:

Grant, J. C., González-Beiras, C., Amick, K. M., Fortney, K. R., Gangaiah, D., Humphreys, T. L., ... Spinola, S. M. (2018). Multiple class I and class II *Haemophilus ducreyi* strains cause cutaneous ulcers in children on an endemic island. *Clinical Infectious Diseases*. <https://doi.org/10.1093/cid/ciy343>

ABSTRACT

Background. Together with *Treponema pallidum* ssp. *pertenue*, *Haemophilus ducreyi* is a major cause of exudative cutaneous ulcers (CU) in children. For *H. ducreyi*, both class I and class II strains, asymptomatic colonization, and environmental reservoirs have been found in endemic regions, but the epidemiology of this infection is unknown.

Methods. Based on published whole genome sequences of *H. ducreyi* CU strains, a single locus typing system was developed and applied to *H. ducreyi* positive CU samples obtained prior to, one, and two years after the initiation of a mass drug administration campaign to eliminate CU on Lihir Island in Papua New Guinea. DNA from the CU samples was amplified with class I and class II *dsrA*-specific primers and sequenced; the samples were classified into *dsrA* types, which were geospatially mapped. Selection pressure analysis was performed on the *dsrA* sequences.

Results. Thirty-seven samples contained class I sequences, 27 contained class II sequences, and 13 contained both. There were 5 class I and 4 class II types circulating on the island; three types accounted for ~87% of the strains. The composition and geospatial distribution of the types varied little over time and there was no evidence of selection pressure.

Conclusions. Multiple strains of *H. ducreyi* cause CU on an endemic island and co-infections are common. In contrast recent findings with *T. p. pertenue*, strain composition is not affected by antibiotic pressure, consistent with environmental reservoirs of *H. ducreyi*. Such reservoirs must be addressed to achieve elimination of *H. ducreyi*.

Key words: *Haemophilus ducreyi*; cutaneous ulcers; single locus typing, molecular epidemiology

In endemic regions of Africa and the South Pacific, 5 to 15% of children have disfiguring exudative cutaneous ulcers (CU). Approximately 100 million children are at risk for CU, and 100,000 cases are reported annually to the World Health Organization (WHO) [1, 2]. Although CU are usually attributed to *Treponema pallidum* ssp. *pertenue*, or yaws, molecular testing shows that *Haemophilus ducreyi* is also a major cause of CU [3-8].

A single oral dose of azithromycin (Az) is effective in the treatment of yaws [9] raising the possibility that mass drug administration (MDA) of oral Az could eradicate yaws. As *H. ducreyi* CU strains are susceptible to Az [10, 11] and Az is effective in the treatment of *H. ducreyi*-associated CU [12], MDA could also eradicate *H. ducreyi*. In the WHO yaws eradication program, MDA is followed every 6 months by case-finding, testing of ulcer swabs, and treatment of new CU cases and household contacts with Az [3]. Following MDA on Lihir Island in Papua New Guinea, the overall prevalence of CU in children fell from ~10% to 1.3% but remained at 1.6% after 42 months of follow-up [13, 14]. The overall number of CU cases with *T. p. pertenue* DNA declined but then significantly increased between 18 and 42 months [14]. Although eight *T. p. pertenue* genotypes were detected, by 24 months only one genotype remained, suggesting the program reduced strain diversity by interrupting transmission [14, 15]. Throughout the study, the percentage of ulcers due to *H. ducreyi* was stable [14]. The failure to eradicate *H. ducreyi* may be due to colonization of the skin of asymptomatic children, flies and bed linens, which allows escape from antimicrobial pressure [16]. In contrast, asymptomatic colonization of the skin by *T. p. pertenue* is rare [16].

H. ducreyi also causes chancroid, a genital ulcer (GU) disease that occurs in Africa and Asia [17]. GU strains are phylogenetically differentiated into two clades, called class I and class II, which diverged from each other approximately 1.95 million years ago [10, 18-21]. Phylogenetic analysis of the whole genome sequences of CU strains shows that both class I and class II strains of *H. ducreyi* cause CU [10, 11, 22, 23]. A multilocus typing system based on *dsrA*, *hgbA*, and *ncaA* recapitulates the phylogenetic tree based on whole genome sequencing [23]. These data suggest that a multilocus typing system could be used for epidemiologic investigations of CU strains, show that both classes of strains cause CU within an endemic country, and raised the possibility that dual infections with both classes of *H. ducreyi* could occur.

Here we examined which components of the multilocus typing system are required to discriminate between CU strains of *H. ducreyi*. Using this information, we typed *H. ducreyi* positive CU samples obtained prior to, one year, and two years after MDA on Lihir Island. We examined whether class I and class II CU strains circulate on the island and whether dual infections occurred. Finally, due to the presence of environmental reservoirs, we tested the hypothesis that the yaws eradication program did not exert selection pressure and/or change the composition or distribution of circulating *H. ducreyi* strains over time.

METHODS

Examination of *dsrA*, *hgbA*, and *ncaA* for polymorphisms among sequenced CU strains

The sequences of *hgbA*, *ncaA*, and *dsrA* were trimmed from the complete genomes of the class I reference strain 35000HP (GenBank accession no. NC_002940.2) and the class II reference strain CIP542 (GenBank accession no. CP011229). The Basic Local Alignment Search Tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to align the gene sequences to each of the 17 previously-sequenced CU strain genomes (GenBank accession numbers CP011218 - CP011221, CP011227, NZ_LMZZ01000014.1, and CP015424 - CP015434). The aligned sequences were analyzed for single nucleotide polymorphisms (SNPs). For *hgbA* and *ncaA*, we compared the entire gene sequences. As there is little homology in *dsrA* between class 1 and class 2 strains, we examined regions of *dsrA* that are amplifiable with the class 1 and class 2-specific primers described below. For class I strains, the primers produce an amplicon that is 441 base pairs (bp) in length. Due to the presence of an unstable repeat region at the 3' end of this amplicon, we restricted our analysis to 363 bp (positions 149-511, which correspond to complete codons 50-169) of the 35000HP reference strain [24]. For class II strains, the primers produce a 711 bp amplicon. The analysis was restricted to bp 78-745 and complete codons 26-247 of the CIP542 reference strain. Manual inspection of the alignments was performed to determine which of the alleles distinguished between the CU strains.

Ethics Statement

All participants, or their parents or guardians, provided oral informed consent to be screened and treated during the yaws eradication campaign; written informed consent was obtained from all patients with lesions before enrollment [3, 13]. The protocol was approved by the National Medical Research Advisory Committee of the Papua New Guinea Ministry of Health (MRAC no. 12.36) [3, 13].

Samples

Swabs of CU were taken before (April 2013), 1 year (May 2014) and 2 years (May 2015) after MDA, as described [3]. The swabs were placed in a lysis / transport buffer that stabilizes DNA. Samples obtained in 2013 were sent to the Molecular Diagnostic Unit at Queensland Royal Brisbane and Women's Hospital (QLD, Australia) for PCR detection of *H. ducreyi* DNA and to the University of Washington (Seattle, WA, USA) for detection of *T. p. pertenue* DNA [3]. Samples obtained in 2014 and 2015 were tested for both *H. ducreyi* and *T. p. pertenue* DNAs at the University Washington. The samples were classified as being positive for *H. ducreyi* DNA, for *T. p. pertenue* DNA, for both, or for neither DNA. Of 272 ulcer samples, 123 contained *H. ducreyi* DNA and 29 contained both *H. ducreyi* and *T. p. pertenue* DNA. Since most of the dual positive samples were exhausted due other analyses, they were excluded from this study. Of the 123 samples, 117 were recovered and used for *H. ducreyi* typing (Figure 1). DNA was stored at -80°C until PCR was performed.

DNA amplification

All specimens underwent PCR with primers specific for class I *dsrA*, based on the sequence of strain 35000HP, using the primers: sense, 5'AGGGTAAATGGACTTGGTCTAATG3'; antisense 5'TGGCTAAACCAGTTTGCAATTC3'. 10-15 µL of each specimen was used as a template in the PCR reaction, which was performed with the Roche Fast Start High Fidelity System. The PCR conditions were as follows: initial denaturation at 94°C for 4 min, followed by 35 cycles of 94 °C denaturation for 1 min, 53° annealing for 1 min, 72 °C elongation for 1 min and final extension for 7 min. To reduce the likelihood of contamination, experiments were conducted under PCR-clean conditions; pipettes,

tips, tubes, and buffers were UV irradiated for 15 minutes. For each batch of specimens amplified, a negative control with no DNA was processed in parallel. Bidirectional sequencing of each PCR product was performed by Eurofins Genomics. If sequencing yielded null or uninterpretable data, a second round of PCR, using conditions identical to those described above, was performed on the first-round PCR products. Negative controls from the first round were also amplified in the second round. If, after the second round of amplification, the sequencing produced null or ambiguous data, the specimen was declared to be negative for class I *H. ducreyi*.

After class I *dsrA* amplification and sequencing had been completed on all samples, the specimens underwent PCR for class II *dsrA*. Class II *dsrA* primers, based on the published sequence of strain CIP542, were sense, 5'GGCATCAAACGGCTCTTTATC3'; antisense, 5'GCTAACGCACTCTTACCTCTAT3'. Negative controls, PCR conditions, DNA sequencing, and data interpretation were identical to those described for class I strains.

Nucleotide alignment and comparison

We used Sequencher 5.4.6 (Gene Codes Corporation, Ann Arbor, MI) to align each ulcer specimen *dsrA* sequence with its respective *dsrA* reference sequence. 35000HP and CIP542 respectively served as the class I and class II reference strains. Differences in *dsrA* sequences among the reference and sequenced CU strains were used to classify each specimen into a "*dsrA* type." Sequences were submitted to GenBank under accession numbers MG953427 - MG953516.

Geospatial mapping

Lihir Island maps were obtained from the Newcrest Mining Department of Community Relations. Three maps of strain distribution were created for the time points before, 1 year, and 2 years after MDA. Village information for each specimen was recorded during collection. Samples with multiple detected strains were mapped as multiple nodes for each strain present. Nodes were manually placed upon the maps at the village locations for each strain.

Selection Pressure analysis

Given the large genetic distance between the two *H. ducreyi* classes, we generated multiple alignments for each collection time for each *dsrA* class type using Sequencher 5.4.6 (Gene Codes Corporation, Ann Arbor, MI). For example, for the data collected before MDA, we created an alignment including the 7 class I.3 types and the 8 class I.4 types, and another separate alignment including 12 class II.3 types and one class II.1 type. Similar alignments were created for the 1 year and 2 year time periods after MDA. Alignments were trimmed to include only complete codons. The class I sequence alignment for the samples collected before MDA included 15 strains and 360 bp (120 codons). The class I sequence alignment for the samples collected 1 year and 2 years after MDA were each 363 bp (121 codons) with 19 and 16 strains, respectively. The class II alignments included 666 bp (222 codons) with 13 samples in the before MDA dataset, 20 samples in the 1 year after MDA dataset, and 7 samples for the 2 year after MDA dataset.

Site-specific selection pressure analysis was performed using five algorithms: fixed effects likelihood (FEL) [25], fast unconstrained Bayesian approximation (FUBAR) [26], internal fixed effects likelihood (IFEL) [27], mixed-effects model of episodic selection (MEME) [28] and the single likelihood ancestor counting (SLAC) [25] using the HyPhy software package [23] as implemented through the Datamonkey server (www.datamonkey.org). The universal genetic code was used for all datasets. The Model Selection tool was used to identify the best fit model for each dataset. The significance level was set at $p < 0.05$ for the FEL, IFEL, MEME, and SLAC analyses; a posterior probability of > 0.9 was used for FUBAR. A site was considered to be under selection if it was identified as such by more than one of the algorithms [29-31]. This was done to account for the uncertainty in whether random-effects models or counting methods are more reliable for detecting selection [32-34].

RESULTS

Development of a single locus typing system

We had shown that a multilocus typing system based on *dsrA*, *ncaA*, and *hgbA* yielded a phylogenetic tree of class I and class II GU and CU strains that was similar to that obtained by whole genome sequencing [21, 23]. We examined which of these alleles or portions of these alleles was required to discriminate among the 17 previously sequenced CU strains and the reference strains 35000HP and CIP542.

Within the 14 previously published class I sequenced CU strains, there were 3 SNPs within *ncaA* and no SNPs within *hgbA* that differed among the CU strains and 35000HP. Within the 363 bp region of *dsrA*, 19 single nucleotide polymorphisms (SNPs) and a variable region corresponding to positions 269-285 of 35000HP were documented (Supplemental Table 1). Within the class I CU strains and 35000HP, variation in *dsrA* alone segregated the strains into 7 types (Supplemental Table 1). Strain 35000HP was unique and was designated type I.1; GHA8 and GHA9 had identical sequences and were designated type I.2; AUSPNG1 was unique and called I.3; NZS4, NZV1, VAN 1, VAN3, VAN4, VAN5 were identical and designated I.4; NZS2 and NZS3 were identical and called I.5; NZS1 was unique and designated I.6; GHA3 and GHA5 were identical and called I.7. The *ncaA* SNPs did not add any discrimination to *dsrA* in segregating these strains (data not shown).

Within the 3 previously published class II CU strains, there were 9 SNPs in *dsrA* (Supplemental Table 2), 2 SNPs in *ncaA*, and 2 SNPs in *hgbA* that distinguished the class II CU strains from CIP542. Strain CIP542 had a unique *dsrA* sequence and was designated as type II.1; GHA1 and GHA2 had the same *dsrA* sequences and were called II.2; VAN2 was unique and called II.3. The *ncaA* SNPs did distinguish between GHA1 and GHA2, but the *hgbA* SNPs did not add any discrimination to *dsrA* in classifying these strains (data not shown). Given that variations in *ncaA* and *hgbA* added little to variations in *dsrA* in distinguishing among the CU strains, we used *dsrA* as a cost efficient single locus typing system for the *H. ducreyi* positive samples from Lihir island.

Sample Typing

Of the 117 *H. ducreyi* positive samples, 77 yielded 90 *dsrA* sequences that could be unambiguously aligned with class I, class II, or both classes of CU strains (Figure 1). Of the 77 specimens, 24 were obtained prior to MDA; 33 were obtained 1 year and 20 were obtained 2 years after MDA. Thirty-seven samples yielded class I amplicons, 27

yielded class II amplicons and 13 yielded both (Figure 1). Polymorphisms found between bp 149 and 511 were used to type the class I samples; polymorphisms found between bp 79 to 744 were used to type the class II samples (Table 1). Overall, there were 50 class I strains comprised of 5 types including one new type (I.8) not present in the previously sequenced CU strains (Figure 1 and Table 1). Within the 40 strains identified as class II, there were 4 types, 2 of which were not present in the previously sequenced strains (Figure 1 and Table 1).

Strain Composition and geospatial mapping

Prior to MDA, there were two class I *dsrA* types (I.3 and I.4) and one class II type (II.3) that were predominant on the island (Figure 2). These types remained predominant throughout the 24 month period and accounted for ~ 18%, 29%, and 40% of the 90 types detected, respectively (Figure 2). Another class II type (II.1) was detected in one sample prior to and 1 year after MDA but was not detected thereafter (Figure 2). In a small number of samples, Types I.1, I.8, and II.4 were detected only 1 year after MDA, while types I.5 and II.5 were detected only 2 years after MDA. Geospatial mapping showed that the three predominant types (I.3, I.4, and II.3) were dispersed over the island throughout the study period (Figure 3).

Selection Pressure Analysis

F81 was found to be the best model for all datasets except the class I samples collected one year after MDA; HKY85 was most appropriate for this dataset. Selection was detected in two of the six datasets by the FUBAR algorithm only, thus no sites were considered to be under selection pressure.

DISCUSSION

We found that single locus typing scheme is sufficient to discriminate among CU strains, that at least nine different types of *H. ducreyi* have been present on Lihir island, and that co-infections with both classes are common, occurring in ~ 17% of 77 cases. Direct whole genome sequencing of 21 CU samples also shows that class I and II strains circulate in Ghana and the Solomon Islands, but co-infections were not detected in this smaller study [35].

Prior to MDA, three types (I.3, I.4, II.3) were predominant on the island. These strains persisted throughout the observation period. Geospatial mapping showed that the predominant strains were dispersed over the entire island throughout the 2 year time period. One type (II.1) was present prior to MDA and one year later, while five types (I.1, I.5, I.8, II.4 and II.5) were detected at a single time point after MDA. As only a convenience sample (90 of 690 CU cases) was tested at baseline [3, 13] and samples that were positive for both *H. ducreyi* and *T. p. pertenue* DNA were excluded from this study, the types that were detected only after MDA could have been present in low levels prior to MDA. Alternatively, the types that were detected only after MDA could have been introduced onto Lihir Island by travel of persons from other endemic islands that were not part of the clinical trial. Overall, ~ 87% of the types detected corresponded to strains that were isolated from the South Pacific previously; no types corresponding to Ghanaian strains were detected.

Limitations of the study include that fact that *dsrA* sequences were found in only 66% of the 117 *H. ducreyi* positive samples; repeated freeze thawing of the samples likely limited our ability to recover unambiguous sequences. As we defined strain types by variations in a single gene, there may be more variability in the circulating strains than what we detected.

Selection pressure analyses indicate that MDA of Az and subsequent treatment of CU cases with Az every 6 months exerted slight or no changes in the composition of the *H. ducreyi* strains during the 2 year observation period. These negative results can be attributed to several reasons: 2 years is a short time frame for selection pressure to become evident in a slowly evolving organism; biannual treatment of patients with CU and their contacts with Az results in very limited pressure on the organism; for selection pressure analysis, our data set was

relatively small. Az is predominantly concentrated intracellularly in fibroblasts [36]; colonizing strains that are on the surface of the skin are likely to escape Az pressure and cause new infections. Finally, given its mechanism of action, Az may not exert pressure on the *dsrA* gene, whose gene product confers serum resistance and binding to fibrinogen, vitronectin, and fibronectin to the organism [37, 38].

In summary, multiple strains of *H. ducreyi* cause CU in an isolated endemic community and co-infections are common. In contrast to the reduction of *T. p. pertenue* diversity [14], MDA did not exert selection pressure and/or change the composition or distribution of circulating *H. ducreyi* strains over a 2-year observation period, consistent with asymptomatic colonization and other environmental reservoirs, which must be addressed to achieve eradication of the *H. ducreyi* component of CU.

Acknowledgments

We thank Sheila Lukehart and Jennifer Robson for providing us with DNA from the clinical samples.

Funding

This work was supported in part by Newcrest Mining Company and by the National Institutes of Health (grant T35 HL110854 to JCG).

Conflicts of interest: All authors: No reported conflicts

REFERENCES

1. Giacani L, Lukehart SA. The endemic treponematoses. *Clinical microbiology reviews* **2014**; 27:89-115.
2. Kazadi WM, Asiedu KB, Agana N, Mitja O. Epidemiology of yaws: an update. *Clinical epidemiology* **2014**; 6:119-28.
3. Mitjà O, Lukehart SA, Pokowas G, et al. *Haemophilus ducreyi* as a cause of skin ulcers in children from a yaws-endemic area of Papua New Guinea: a prospective cohort study. *The Lancet Global Health* **2014**; 2:e235-e41.
4. Marks M, Chi KH, Vahi V, et al. *Haemophilus ducreyi* Associated with Skin Ulcers among Children, Solomon Islands. *Emerging infectious diseases* **2014**; 20:1705-7.
5. Ghinai R, El-Duah P, Chi KH, et al. A cross-sectional study of 'yaws' in districts of Ghana which have previously undertaken azithromycin mass drug administration for trachoma control. *PLoS neglected tropical diseases* **2015**; 9:e0003496.
6. Chi KH, Danavall D, Taleo F, et al. Molecular differentiation of *Treponema pallidum* subspecies in skin ulceration clinically suspected as yaws in Vanuatu using real-time multiplex PCR and serological methods. *The American journal of tropical medicine and hygiene* **2015**; 92:134-8.
7. Fegan D, Glennon MJ, Kool J, Taleo F. Tropical leg ulcers in children: more than yaws. *Tropical doctor* **2016**; 46:90-3.
8. Gonzalez-Beiras C, Marks M, Chen CY, Roberts S, Mitja O. Epidemiology of *Haemophilus ducreyi* Infections. *Emerging infectious diseases* **2016**; 22.

9. Mitja O, Hays R, Ipai A, et al. Single-dose azithromycin versus benzathine benzylpenicillin for treatment of yaws in children in Papua New Guinea: an open-label, non-inferiority, randomised trial. *Lancet* **2012**; 379:342-7.
10. Gangaiah D, Webb KM, Humphreys TL, et al. *Haemophilus ducreyi* Cutaneous Ulcer Strains Are Nearly Identical to Class I Genital Ulcer Strains. *PLoS neglected tropical diseases* **2015**; 9:e0003918.
11. Gangaiah D, Marinov GK, Roberts SA, Robson J, Spinola SM. Draft Whole-Genome Sequence of *Haemophilus ducreyi* Strain AUSPNG1, Isolated from a Cutaneous Ulcer of a Child from Papua New Guinea. *Genome announcements* **2016**; 4.
12. Gonzalez-Beiras C, Kapa A, Vall-Mayans M, et al. Single-Dose Azithromycin for the Treatment of *Haemophilus ducreyi* Skin Ulcers in Papua New Guinea. *Clinical Infectious Diseases* **2017**; 65:2085-90.
13. Mitja O, Houinei W, Moses P, et al. Mass treatment with single-dose azithromycin for yaws. *The New England journal of medicine* **2015**; 372:703-10.
14. Mitja O, Godornes C, Houinei W, et al. Re-emergence of yaws after single mass azithromycin treatment followed by targeted treatment: a longitudinal study. *Lancet* **2018**, *in press*.
15. Godornes C, Giacani L, Barry AE, Mitja O, Lukehart SA. Development of a Multilocus Sequence Typing (MLST) scheme for *Treponema pallidum* subsp. *pertenue*: Application to yaws in Lihir Island, Papua New Guinea. *PLoS neglected tropical diseases* **2017**; 11:e0006113.
16. Houinei W, Godornes C, Kapa A, et al. *Haemophilus ducreyi* DNA is detectable on the skin of asymptomatic children, flies and fomites in villages of Papua New Guinea. *PLoS neglected tropical diseases* **2017**; 11:e0004958.

17. Spinola SM. Chancroid and *Haemophilus ducreyi*. In: Holmes KK, Sparling PF, Stamm WE, et al., eds. Sexually transmitted diseases. 4th ed. New York: McGraw-Hill, **2008**:689-99.
18. White CD, Leduc I, Jeter C, Harris C, Elkins C. *Haemophilus ducreyi* outer membrane determinants, including DsrA, define two clonal populations. Infect Immun **2005**; 73:2387-99.
19. Post DM, Gibson BW. Proposed second class of *Haemophilus ducreyi* strains show altered protein and lipooligosaccharide profiles. Proteomics **2007**; 7:3131-42.
20. Post DMB, Munson RS, Jr., Baker B, Zhong H, Bozue JA, Gibson BW. Identification of genes involved in the expression of atypical lipooligosaccharide structures from a second class of *Haemophilus ducreyi*. Infect Immun **2007**; 75:113-21.
21. Ricotta EE, Wang N, Cutler R, Lawrence JG, Humphreys TL. Rapid Divergence of Two Classes of *Haemophilus ducreyi*. Journal of bacteriology **2011**; 193:2941-7.
22. Pillay A, Katz SS, Abrams AJ, et al. Complete Genome Sequences of 11 *Haemophilus ducreyi* Isolates from Children with Cutaneous Lesions in Vanuatu and Ghana. Genome announcements **2016**; 4.
23. Gangaiah D, Spinola SM. *Haemophilus ducreyi* Cutaneous Ulcer Strains Diverged from Both Class I and Class II Genital Ulcer Strains: Implications for Epidemiological Studies. PLoS neglected tropical diseases **2016**; 10:e0005259.
24. Elkins C, Morrow KJ, Olsen B. Serum resistance in *Haemophilus ducreyi* requires outer membrane protein *DsrA*. Infect Immun **2000**; 68:1608-19.
25. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics **2005**; 21:676-9.

26. Murrell B, Moola S, Mabona A, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular biology and evolution* **2013**; 30:1196-205.
27. Pond SL, Frost SD, Grossman Z, Gravenor MB, Richman DD, Brown AJ. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comput Biol* **2006**; 2:e62.
28. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS genetics* **2012**; 8:e1002764.
29. Lamers SL, Newman RM, Laeyendecker O, et al. Global Diversity within and between Human Herpesvirus 1 and 2 Glycoproteins. *Journal of virology* **2015**; 89:8206-18.
30. Minaya MA, Korom M, Wang H, Belshe RB, Morrison LA. The herpevac trial for women: Sequence analysis of glycoproteins from viruses obtained from infected subjects. *PloS one* **2017**; 12:e0176687.
31. Priyam M, Tripathy M, Rai U, Ghorai SM. Divergence of protein sensing (TLR 4, 5) and nucleic acid sensing (TLR 3, 7) within the reptilian lineage. *Molecular phylogenetics and evolution* **2018**; 119:210-24.
32. Suzuki Y, Nei M. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. *Molecular biology and evolution* **2001**; 18:2179-85.
33. Suzuki Y, Nei M. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Molecular biology and evolution* **2002**; 19:1865-9.
34. Suzuki Y, Nei M. False-positive selection identified by ML-based methods: examples from the Sig1 gene of the diatom *Thalassiosira weissflogii* and the tax gene of a human T-cell lymphotropic virus. *Molecular biology and evolution* **2004**; 21:914-21.

35. Marks M, Fookes M, Wagner J, et al. Direct Whole-Genome Sequencing of Cutaneous Strains of *Haemophilus ducreyi*. *Emerging infectious diseases* **2018**; 24:786-9.
36. Thornton AC, O'Mara EM, Jr., Sorensen SJ, et al. Prevention of experimental *Haemophilus ducreyi* infection: a randomized, controlled clinical trial. *J Infect Dis* **1998**; 177:1608-13.
37. Leduc I, White CD, Nepluev I, Throm RE, Spinola SM, Elkins C. Outer membrane protein DsrA is the major fibronectin-binding determinant of *Haemophilus ducreyi*. *Infect Immun* **2008**; 76:1608-16.
38. Leduc I, Olsen B, Elkins C. Localization of the domains of the *Haemophilus ducreyi* trimeric autotransporter DsrA involved in serum resistance and binding to the extracellular matrix proteins fibronectin and vitronectin. *Infect Immun* **2009**; 77:657-66.

Table 1. Polymorphisms in *dsrA* used to type the *H. ducreyi*-positive clinical samples.

Class I																				
Type	152	170	175	193	196	209	215	226	227	266	269	271	271.1	272	272.1	272.2	273	274	275	277
I.1	G	T	G	A	G	A	C	G	A	C	A	G	-	C	-	-	T	C	C	G
I.2	A	C	T	A
I.3	A	C	T	A	.	.	C	.	.	A	.	.	C	G	.	C
I.4	A	C	A	.	.	.	T	A	.	.	C	.	.	A	.	.	C	G	T	C
I.5	A	C	A	.	.	.	T	A	.	.	C	.	A	.	T	T	.	.	.	C
I.6	A	C	A	.	.	.	T	A	.	.	C	.	.	-	.	.	A	.	G	-
I.7	.	C	.	.	A	G	.	A	G	A	.	C	T	.	-
I.8*	A	C	A	G	.	.	T	A	.	.	C	.	.	A	.	.	C	G	T	C

Class I (continued)																			
Type	278	279	280	281	282	283	284	285	290	292	299	313	405	411	427	460	469	476	491
I.1	G	C	G	T	T	T	C	T	G	C	T	T	A	T	A	T	G	G	A
I.2	A	.	.	C	.	C	.	.	A	T	.	.	T	G
I.3	C	T	C	C	.	C	.	.	.	T	.	.	T	G
I.4	C	T	C	C	.	C	.	.	.	T	.	.	T	.	.	C	.	.	G
I.5	C	T	C	C	.	C	.	.	.	T	.	.	T	.	.	C	.	.	G
I.6	T	.	-	C	.	C	.	.	.	T	.	.	T	.	.	C	.	.	G
I.7	.	.	-	-	A	A	A	A	.	A	G	C	T	.	C	.	A	A	G
I.8*	C	T	C	C	.	C	.	.	.	T	.	.	T	C	.	C	.	.	G

Class II											
Type	125	214	216	273	319	344	346	358	359	380	383
II.1	G	G	T	T	C	T	A	T	A	G	G
II.2	.	.	G	A	A	.	G	C	G	.	T
II.3	.	A	.	A	A	.	G	C	.	A	T
II.4*	C
II.5*	A	A	.	A	A	.	G	C	.	A	T

Data represent single nucleotide polymorphisms in the *dsrA* amplicons that were used to type the clinical samples. For the class I types, the numbers correspond to the bp of the *dsrA* allele of 35000HP (type I.1); for the class II types, the numbers correspond to the bp of the *dsrA* allele of CIP542 (type II.1). A dot (.) indicates a match to the reference sequence and a dash (-) indicates that that position does not exist in the sample. An asterisk (*) denotes a novel type.

FIGURE LEGENDS

Figure 1. Sample flow chart for *H. ducreyi* positive specimens from Lihir Island and overall results of *dsrA* typing. The *dsrA* types are defined in Table 1; the numbers of positive samples for each type are in parentheses. An asterisk (*) denotes a novel type.

Figure 2. Percentage of *H. ducreyi dsrA* types detected on Lihir Island detected prior to and 1 and 2 years after MDA. For dual infections, both types were included.

Figure 3. Geospatial mapping of *H. ducreyi* strains on Lihir Island before MDA (A), 1 year (B), and 2 years after MDA (C). If a sample contained more than one *dsrA* type, both types were mapped to the same village.

Figure 1.

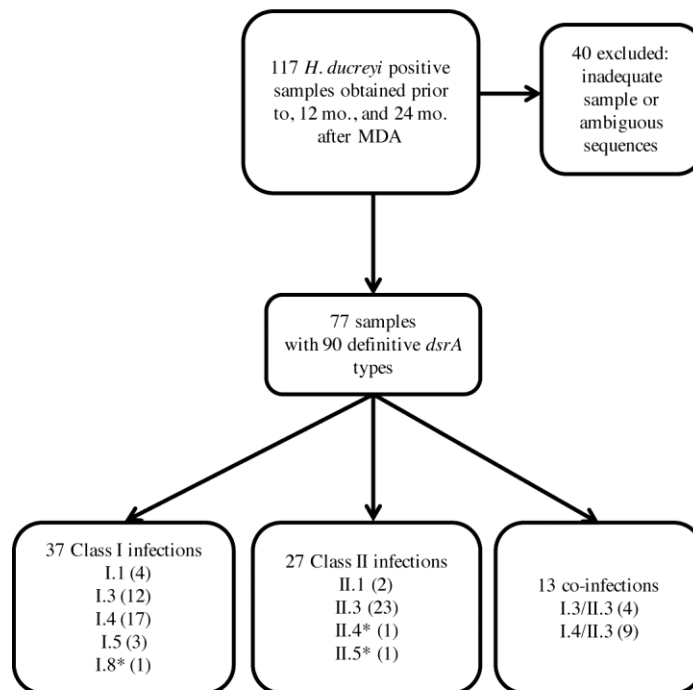


Figure 2.

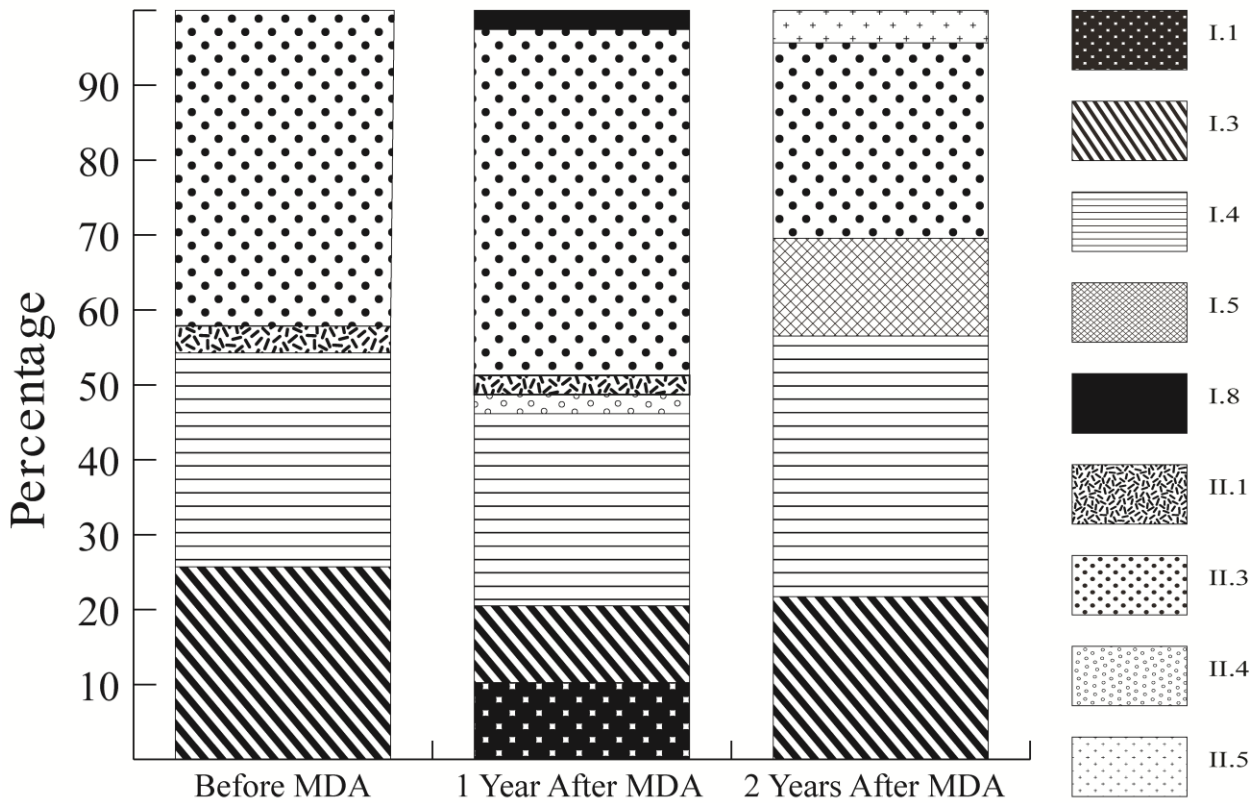


Figure 3.

